DOCUMENT RESUME

ED 298 170 TM 012 319

AUTHOR Plake, Barbara S.; Wise, Steven L.
TITLE Examinee Selection of Subsequent Item Difficulty: Effects of Current Item Performance and Item Feedback.
PUB DATE [86]
NOTE 13p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; Algebra; *College Students; Computer Assisted Testing; *Difficulty Level; *Feedback; Higher Education; *Item Analysis; Latent Trait Theory; Mathematics Tests; Multiple Choice Tests; Response Style (Tests); Test Construction; *Test Format; Test Items

ABSTRACT

One question regarding the utility of adaptive testing is the effect of individualized item arrangements on examinee test scores. The purpose of this study was to analyze the item difficulty choices by examinees as a function of previous item performance. The examination was a 25-item test of basic algebra skills given to 36 students in an introductory statistics course at a large midwestern university during the fall 1985 semester. The test was administered via a microcomputer. Categorical data identifying the subjects' performance on current item, feedback condition, and choice of difficulty for next item were submitted to a log-linear analysis. A significant performance on current item by choice of difficulty for next item interaction was found. When examinees did not answer the current item correctly, they tended to request an easier item next; when they correctly answered the current item, they tended to request a harder item next. Preference for the difficulty of the next item appears to be a function of how well students believe they performed on the current item. Most adaptive testing item selection algorithms identify an easier item to be administered upon incorrect performance and a harder item subsequent to successful item performance. Results of this study suggest that this selection algorithm is congruent with examinee selected item difficulty. A table and a graph present study data. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED298170

Examinee Selection of Subsequent Item Difficulty: Effects of Current Item Performance and Item Feedback

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Barbara S. Plake

Steven L. Wise

University of Nebraska

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BARBARA S. PLAKE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM012 319

Examinee Selection of Subsequent Item Difficulty:
Effects of Current Item Performance and Item Feedback

(Abstract)

One question regarding the utility of adaptive testing is the effect of individualized item arrangements on examinee test scores. The purpose of this study was to analyze the item difficulty choices by examinees as a function of previous item performance. Categorical data identifying subjects' performance on current item, feedback condition, and choice of difficulty for next item were submitted to a log-linear analysis. A significant performance on current item by choice of difficulty for next item interaction was found. When examinees did not answer the current item correctly, they tended to request an easier item next; when they correctly answered the current item, they tended to request a harder item next. Preference for the difficulty of the next item appears to be a function of how well students believe they performed on the current item. Most adaptive testing item selection algorithms identify an easier item to be administered upon incorrect performance and a harder item subsequent to successful item performance. Results of this study suggest that this selection algorithm is congruent with examinee selected item difficulty choices.

Examinee Selection of Subsequent Item Difficulty: Effects of Current Item Performance and Item Feedback

In most multiple-choice examination administrations, examinees proceed through the test in a fixed item arrangement. Often these items are arranged, based on pilot data, from easiest to hardest. Other item arrangements have also been used and studied, including easy-to-hard within content, spiral-cyclical, hard-to-easy, and random. Some evidence for effects of item arrangement on test performance has been found, especially for quantitative content areas (Leary & Dorans, 1985; Plake, Ansorge, Parker, & Lowry, 1982; Wise, Plake, Eastman, Boettcher, & Lukin, 1986).

When a multiple choice examination is administered by a computer, many previously fixed (or difficult to vary) dimensions of testing become more flexible. One such dimension is the individualization of the order of item presentation to the examinee. In adaptive testing, for example, the decision of what items, and in what order of presentation, is based in part on examinee performance on previous items. Given that item context has been shown to affect test performance in quantitative areas, one question regarding the utility of adaptive testing is the effect of individualized item arrangements on examinee test scores.

A recent study by Wise et al. (1986) investigated the effect of variation of individualized item arrangement on test performance in a quantitative content area. Twenty-five mathematics items were administered via computer to examinees.

Three item arrangements were considered: easy-to-hard, random, and examinee selected. In the examinee selected item arrangement, all examinees began with the item of medium difficulty. After answering that item, examinees were asked by the computer if they wanted the next item to be easier or harder. Half of the examinees, regardless of item arrangement condition, received feedback on the correctness of their answers to items. Therefore, in choosing whether to ask for an easier or harder item, half of the examinees in the examinee selected item arrangement condition had information about correctness of their answer to the previous item. In that study, although a significant arrangement by feedback effect on test performance was found, the effect did not involve the examinee selected item arrangement condition.

Therefore, the study by Wise et al. (1986) suggests that individualization of item arrangement by examinee preference does not have an impact of resulting test scores. However, additional questions about examinee preference for item arrangement, and how that preference might vary as a function of feedback, are relevant to individualization of item arrangement. The purpose of this study was to analyze the item difficulty choices by examinees as a function of previous item performance and feedback. Specifically, this study investigated examinee preference of the difficulty of subsequent items under conditions of knowledge or no

knowledge of correctness of performance on current test questions.

Method

Subjects

The subjects were 36 students who had been randomly assigned to the examinee selected item arrangement condition in the larger Wise et al. (1986) study. All of the subjects were students in an multisectioned introductory statistics course at a large midwestern university in the fall, 1985 semester. Testing was done for instructional purposes as a required part of the students' introductory statistics course.

Instrument

The examination was a 25-item test designed to measure basic algebra skills necessary for student understanding of introductory statistics. The test was administered via Apple IIe microcomputers. Prior to this study, the 25-multiple choice algebra test items were administered to a group of introductory statistics students to estimate the difficulty (proportion correct) of each item. The total-test reliability was estimated, using KR_{20}, to be .82.

In the examinee selected item arrangement condition, the student was given some control over the order of difficulty of the presentation of the 25 items. The first item administered was of median difficulty. After answering each item, the student was asked to choose whether he or she wanted to be administered an item more or less difficult than the one just answered. Under

this choice algorithm, however, students might eventually make choices that could not be accommodated due to the fixed item pool. For example, midway through the test a student might choose to be an easier item when there were no more remaining items available that were easier than the previous item. When this happened, the student was informed by the computer that his or her choice could not be met, and he or she was then administered the item of median difficulty from the remaining set of unadministered items. Through use of this algorithm, each student was eventually administered all 25 items in the test. Half of the examinees received feedback regarding the correctness of their answer to each item by the computer.

The data for this study consisted of the first five items administered under the examinee selected item arrangement condition in the Wise et al. (1986) study. Only the first five item selections were used because, as examinees proceeded through the fixed set of items, their ability to have control of the relative difficulty of subsequent items diminished as the pool of remaining items became restricted by their previous item choices. It was decided the five items were sufficient to establish a trend in item choice patterns while still maintaining a high degree of examinee control over relative item difficulty. For each examinee, the data available on the first five items were (a) feedback condition (yes or no), (b) performance on current item (correct or incorrect), and (c) choice of difficulty for next item (easier or harder). These categorical data were

submitted to a log-linear analysis to investigate the relationship between choice of difficulty of next item as a function of performance on current item and feedback condition.

Procedure

On the first day of class students signed up for a time to be administered the algebra test. The students were told they were required to take a test which focused on algebra skills needed to understand introductory statistics. The students were also told that, if they received a "low" score, they would be required to attend a one-hour remediation session during which basic algebra skills would be reviewed. The cut-off score for determining which students would have to attend the remediation session was withheld until all students had completed the test.

The algebra test was administered during the first five days of the fall, 1985 semester. When each student arrived for testing, he or she was seated at an Apple IIe microcomputer and given a brief orientation on how to use the microcomputer for testing. Next, the student proceeded through the computer-administered test.

Results

Table 1 summarizes the frequency counts per condition. Of the main and interaction effects accessible in the loglinear analysis, the following were found to be significant at the $\alpha = .05$ level: Main effect for next item difficulty level (58% chose an easier item next regardless of item performance and/or feedback condition); main effect for item performance (72% got the

current item correct), and interaction between correctness of current item and choice of difficulty of next item (regardless of feedback) (See Figure 1). When examinees did not answer the current item correctly (whether they explicitly knew this from feedback or not), 85% of the time they requested an easier item for the next item. However, when examinees correctly answered the current item (again regardless of feedback condition), 44% of the time they asked for an easier item to be administered next. Since the triple interaction of current item performance, feedback, and next item choice was not significant, the decision of preference of difficulty for the item to be administered next appears to be a function of examinees' self-determined perception of item performance.

Conclusions

Preference for the difficulty of the next item in an examination appears to be a function of how well students believe they performed on the current item. However, this result must be interpreted in the context of the examination situation. Examinees knew that it was not possible to always ask for easier items since the same set of 25 items were to be administered to all examinees sometime in the testing process. Further, no information is present in this analysis of the effect of a history of performance on item difficulty choices. It may be that the choice of difficulty of the subsequent item is related to a

compound effect of several correct (or incorrect) answers in a row.

In most adaptive testing environments, item administration will be individualized based on current item performance. Most item selection algorithms in the adaptive models identify an easier item to be administered upon incorrect item performance and a harder item subsequent to successful item performance. Results from this study suggest that such an item selection strategy is consistent with examinee selected item difficulty choices.

References

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. Review of Educational Research, 55(3), 387-413.

Plake, B. S., Ansorge, C. J., Parker, C. S., & Lowry, S. N. (1982). Effects of item arrangement, knowledge of arrangement, test anxiety, and sex on test performance. Journal of Educational Measurement, 19(1), 49-57.

Wise, S. L., Plake, B. S., Eastman, L. E., Boettcher, L. L., & Lukin, M. E. (1986). The effects of item feedback and examinee control on test performance and anxiety in a computer-administered test. Computers in Human Behavior, 2, 21-29.

Table 1

Observed Frequencies by Condition

Item Feedback?	Response to current item	Requested next item to be easier	Requested next item to be harder
Yes	Correct	27	32
Yes	Incorrect	23	3
No	Correct	33	44
No	Incorrect	23	5

Figure 1. Interaction of EASEHARD by ANSCORE